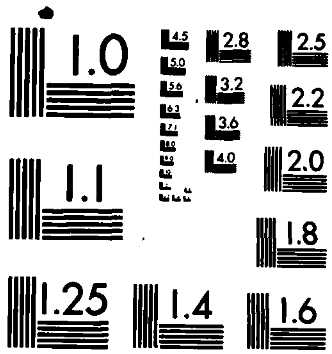
AD-A161 890   DESIGN OF GRAPHIC DISPLAYS IN COMPUTERIZED SYSTEMS(U)   1/1
AMERICAN SOCIETY FOR ENGINEERING EDUCATION WASHINGTON
DC K BENNETT 01 SEP 85 N00014-83-D-6689

UNCLASSIFIED   F/G 9/2   NL

END


1.0
4.5
5.0
5.6
6.3
2.8
3.2
3.6
4.0
2.5
2.2
2.0
1.1
1.8
1.25
1.4
1.6

OFFICE OF NAVAL TECHNOLOGY

POST-DOCTORAL FELLOWSHIP PROGRAM

administered by

THE AMERICAN SOCIETY FOR ENGINEERING EDUCATION

Washington, D. C. 20036

DESIGN OF GRAPHIC DISPLAYS IN COMPUTERIZED SYSTEMS

KEVIN BENNETT

Contract No. N00014-83-D-0689

FINAL REPORT (9/1/84 TO 9/1/85)

Abstract

The user's mental model of a computerized, perceptual database system was investigated in three experiments. The system consisted of a database of multidimensional sounds, commands to search the database, and one of three separate displays (two graphic displays for training, an alpha-numeric display for testing). The graphic displays presented different conceptualizations of the database; training with a different graphic display was predicted to cause the formation of a different mental model of the system. The results of three experiments indicated that users trained with one graphic display identified two-dimensional sounds with significantly lower latency (Experiment 1) than users trained with the second graphic display. For three-dimensional sounds these findings were reversed (Experiment 2). When the user was trained with both displays this interaction disappeared (Experiment 3). The results indicate that display design can influence the user's mental model of a system and that this has implications for performance with the system. Keywords: man machine systems

In the last few decades there has been a dramatic increase in the complexity of man-machine systems. Process control, nuclear submarines, and aeronautics are but a few examples. The increase in complexity of the underlying systems has been accompanied by an increase in the technology available for the man-machine interface. These advances include the development of computerized graphics, voice I/O and artificial intelligence. The result is a need to assess the implications of these technologies for system design.

This assessment is the responsibility of a number of disciplines, most notably experimental psychology, engineering psychology, and human factors. Wickens (1984, pp. 3-4) describes the approach of each discipline in the following statement:

> "The goal of experimental psychology is to uncover the laws of behavior through experiments. However, the design of these experiments is unconstrained by a requirement to apply the laws. That is, it is not required that experiments generate immediately useful information. The goal of human factors, on the other hand, is to apply knowledge in designing systems that work, accommodating the limits of human performance and

exploiting the advantages of the human operator in the process. Engineering psychology arises from the convergence of these two domains. 'The aim of engineering psychology is not simply to compare two possible designs for a piece of equipment [which is the role of human factors], but to specify the capacities and limitations of the human [generate an experimental database] from which the choice of a better design should be directly deducible' (Poulton, 1966, p.178). That is, while research topics in engineering psychology are selected because of applied needs, the research transcends specific one-time applications and is conducted with the broader objective of providing a usable theory of human performance."

The work that has been completed for the ONT Postdoctoral Fellowship falls under the category of engineering psychology. The experiments were conducted with the goal of obtaining generalizable data on visual reasoning and imaging in a systems context. Specifically, the research concentrated on the interaction between the design of graphic displays and the information-processing capabilities of the user. The research was motivated by several issues which are important for the design of complex systems.

First, computer graphics are becoming an increasingly important aspect of the man-machine interface. Computer graphics are an extremely efficient method for data presentation. A number of theories have been advanced to explain this phenomena, but let it suffice that graphic presentation allows the interrelations among data to be easily seen and integrated. Also, the use of software-generated controls and displays allows great flexibility in the design of machines. Examples can be seen in the design of programming environments (Glinert & Tanimoto, 1984) and operating systems (Apple's <u>Lisa</u> and Xerox's <u>Star</u>) in which all interaction is handled graphically.

The second factor motivating the research is an interest in the information-processing capabilities of the user and how this is affected by system design. If a computing system is to be optimally effective the user's capabilities and limitations must be considered in its design. Traditionally, human factors has focused on design constraints imposed by the physical characteristics of the user (e.g., angle of the keyboard or VDT). However, human-computer interaction is an activity that is highly knowledge-intensive. This dictates the study of a less tractible but potentially more rewarding domain: user cognition. Hollnagel and Woods (1983) describe a perspective to system design, which they have termed cognitive systems engineering (CSE), which incorporates this philosophy. They state:

> "In contrast to traditional approaches to the study of man-machine systems which mainly operate on the physical and physiological level, CSE operates on the level of cognitive functions. Instead of viewing an MMS as decomposable by mechanistic principles, CSE introduces the concept of a cognitive system: an adaptive system which functions using knowledge about itself and the environment in the planning and modification of actions. Operators are generally acknowledged to use a model of the system (machine) with which they work. Similarly, the machine has an image of the operator. The designer of an MMS must recognize this, and strive to obtain a match between the machine's image and the user characteristics on a cognitive level, rather than just on the level of physical functions."

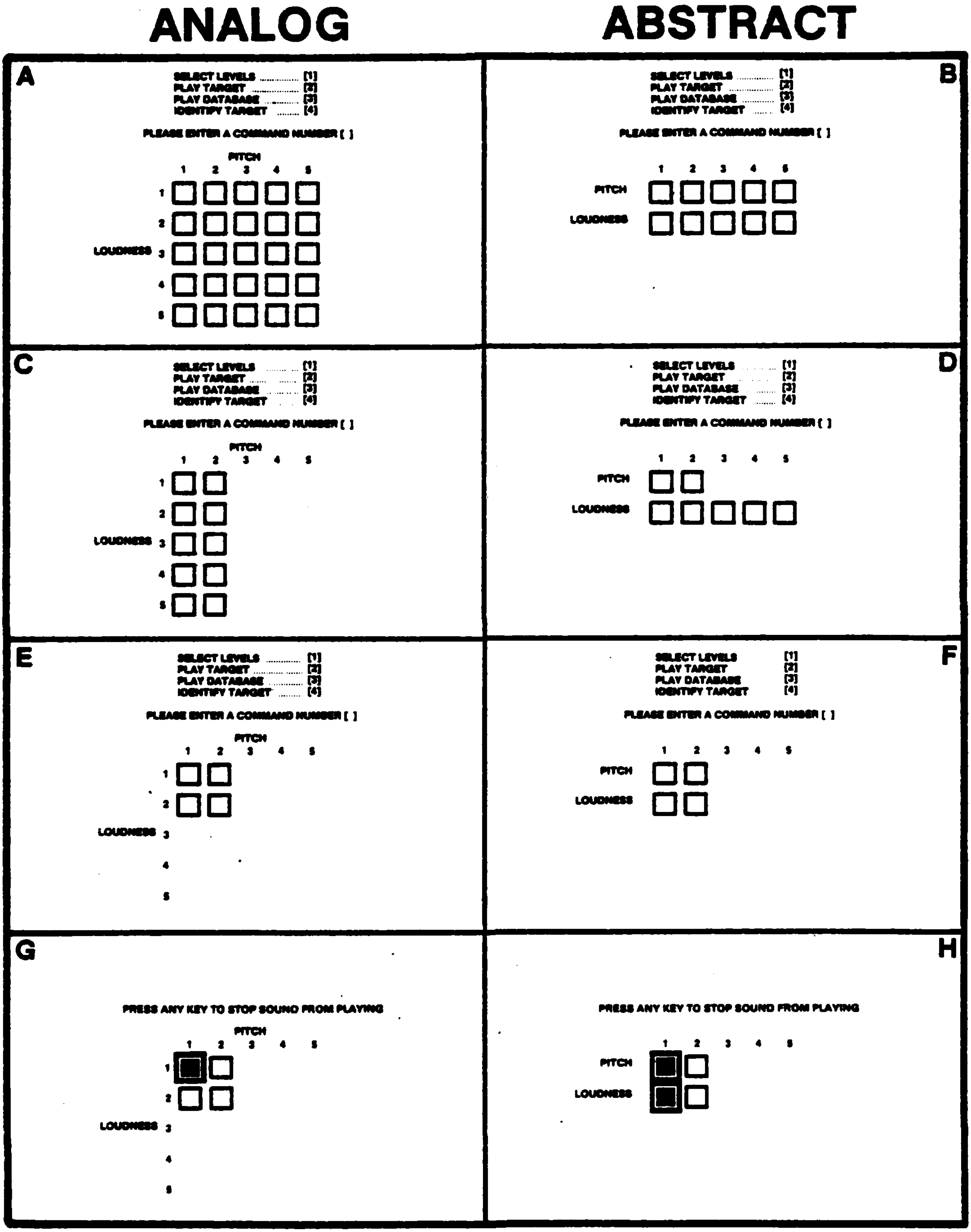
The present research is concerned with the model of the system that users have. This has alternately been referred to as the user's analogical (Rumelhart and Norman, 1981), metaphorical (Carroll and Thomas, 1982), and qualitative (Williams, Hollan, and Stevens, 1981) reasoning, the user's conceptual model (Young, 1981, 1983), and the term that the present paper will adopt: the user's mental model

(Carey, 1982; Halasz and Moran, 1982, 1983; Hollan, Hutchins, and Weitzman, 1984; Moran, 1981a, 1981b; Norman, 1983).

Young (1981) issues a qualified definition, stating that the concept of the user's mental model of an interactive device "is a rather hazy one, but central to it is the assumption that the user will adopt some more or less definite representation or metaphor which guides his actions and helps him interpret the device's behavior" (p.51). The user's mental model of an interactive computer system includes knowledge of the internal workings of the system, what tasks can be accomplished with the system, and how to accomplish those tasks. Essentially, the user's mental model is the knowledge (and/or beliefs) about a system that an individual uses to operate the system.

Previous articles have discussed mental models of programming languages (du Boulay, O'Shea, and Monk, 1981; Mayer, 1980, 1981), calculators (Halasz and Moran, 1983; Young, 1981, 1983), and complex systems (Carey, 1982; Hollan, Hutchins, and Weitzman, 1984; Moran, 1981a, 1981b; Williams, Hollan, and Stevens, 1981). In general, it is claimed that the user's mental model of an interactive device is influenced by information from a variety of sources including the design of training materials, system manuals, and system interface. If these components are well designed and complementary then the user is likely to form an appropriate mental model of the system. However, the vast majority of these articles are not empirical in nature. If the user's mental model is to be a consideration in the design of computing systems there must be empirical evidence indicating that design can influence the user's mental model of a system and that this has implications for performance with the system.

=========================


ANALOG
ABSTRACT
A
SELECT LEVELS [1]
PLAY TARGET [2]
PLAY DATABASE [3]
IDENTIFY TARGET [4]
PLEASE ENTER A COMMAND NUMBER [ ]
PITCH
1 2 3 4 5
LOUDNESS
1 2 3 4 5
B
SELECT LEVELS [1]
PLAY TARGET [2]
PLAY DATABASE [3]
IDENTIFY TARGET [4]
PLEASE ENTER A COMMAND NUMBER [ ]
1 2 3 4 5
PITCH
LOUDNESS
C
SELECT LEVELS [1]
PLAY TARGET [2]
PLAY DATABASE [3]
IDENTIFY TARGET [4]
PLEASE ENTER A COMMAND NUMBER [ ]
PITCH
1 2 3 4 5
LOUDNESS
1 2 3 4 5
D
SELECT LEVELS [1]
PLAY TARGET [2]
PLAY DATABASE [3]
IDENTIFY TARGET [4]
PLEASE ENTER A COMMAND NUMBER [ ]
1 2 3 4 5
PITCH
LOUDNESS
E
SELECT LEVELS [1]
PLAY TARGET [2]
PLAY DATABASE [3]
IDENTIFY TARGET [4]
PLEASE ENTER A COMMAND NUMBER [ ]
PITCH
1 2 3 4 5
LOUDNESS
1 2 3 4 5
F
SELECT LEVELS [1]
PLAY TARGET [2]
PLAY DATABASE [3]
IDENTIFY TARGET [4]
PLEASE ENTER A COMMAND NUMBER [ ]
1 2 3 4 5
PITCH
LOUDNESS
G
PRESS ANY KEY TO STOP SOUND FROM PLAYING
PITCH
1 2 3 4 5
LOUDNESS
1 2 3 4 5
H
PRESS ANY KEY TO STOP SOUND FROM PLAYING
1 2 3 4 5
PITCH
LOUDNESS

Insert Figure 1 about here

========================

To address these issues a computerized, perceptual database system was developed that had multidimensional sounds as its database. Three separate interfaces to the system were designed which differed only in the displays used to present information. Two displays were graphic and represented different conceptualizations of the database, as shown in Figure 1. The display on the left (Figure 1, parts a,c,e, and g) is referred to as the analog display because of the direct correspondence between a sound in the database and a box in the display; the display on the right (Figure 1, parts b,d,f, and h) is referred to as the abstract display because it does not have this one-to-one correspondence. A third display, the alpha-numeric display, presented information with numbers and letters rather than graphics (boxes). A detailed explanation of the database system and system displays is given in the methods section of Experiment 1.

Training with a different graphic display was hypothesized to result in a different mental model of the system. Three experiments were conducted to test this hypothesis. In the experiments users were trained to search a two-dimensional database (sounds varying in pitch and loudness) with the analog and/or the abstract graphic display(s) and were then tested with the alpha-numeric display. In the first experiment users identified both two-dimensional and three-dimensional sounds (varying in pitch, loudness, and duration) during testing (Days 2 and 3).

EXPERIMENT 1

It was hypothesized that training with a graphic display would provide users with an internal model of the system for reasoning about

the identification of sounds. When tested with the interface containing the alpha-numeric display users would reason about the task in terms of the graphic display that they had been trained with. Performance differences were expected because each graphic display was more or less appropriate for the two- or the three-dimensional database.

First, consider the analog display (see Figure 1, part a). This display was a particularly appropriate representation of the two-dimensional database because of the one-to-one relation between a box in the display and a sound in the database. The direct correspondence produced a display that was very spatial in nature: the display represented a sound by one box in a particular area. The abstract display lacked this direct correspondence. Therefore, training with the analog display should provide users with a more appropriate mental model of the database system for the identification of two-dimensional sounds.

For three-dimensional sounds the situation is reversed with the abstract display providing a more appropriate representation. Remember that the graphic displays were only two-dimensional in nature: users would have to extend the display they were trained with to represent a three-dimensional database. The abstract display could be easily extended to represent a database of any dimensionality by adding a row of boxes for an additional dimension of sound. On the other hand, the analog display could not be easily extended. It would be necessary to imagine a cube (or three planes) to represent a three-dimensional database. Thus, training with the abstract display should provide users with a more appropriate mental model for the identification of three-dimensional sounds than the analog display.

Users trained with the abstract display were predicted to identify three-dimensional sounds with lower latency and higher accuracy than users trained with the analog display. Users trained with the analog display were predicted to identify two-dimensional sounds with lower latency and higher accuracy than users trained with the abstract display.

METHOD

Subjects. Forty-four volunteers from an introductory psychology class, aged 18 to 22 years, participated for credit. Four subjects were dropped from the analysis due to a failure to complete the task. No listeners reported a history of hearing disorders. Subjects were assigned randomly to graphic display (analog or abstract) and to order of testing (two- or three-dimensional sounds first).

Stimuli. All sounds were synthesized on a digital computer using standard algorithms. The two-dimensional database contained 25 sounds constructed by a factorial combination of five levels of pitch (920, 978, 1040, 1105, and 1175 Hz) and five levels of loudness (75, 78, 81, 84, and 87 dB SPL). The three-dimensional database contained 125 sounds constructed by a factorial combination of the pitch, the loudness, and five duration levels (100, 220, 340, 460, and 580 msec).

The database system had five commands: 1) select levels, 2) play target, 3) play database, 4) identify target, and 5) select order. Figure 1 illustrates both graphic display in response to several of these commands. Figure 1 (parts a and b) illustrates each graphic display at the beginning of a trial. To aid in identification the user could select a subset of the database to compare to the target sound. Figure 1 (parts c and d) illustrates each display in response to the select levels command which was used to decrease the range of

pitch levels. Figure 1 (parts e and f) show the response of each display to a similar command for loudness. At this point the database would contain four sounds and the play database command would play these four sounds. As each sound was played the box (analog) or boxes (abstract) representing that sound was placed in reverse graphics (black on green, instead of green on black). Figure 1 (parts g and h) illustrate each display as it appeared when the first sound in the selected portion of the database was played. The select order command was used [illegible] that these sounds were played. The play target command was used to play the target sound. By reducing the number of sounds the user could compare successively smaller and smaller portions of the database to the target sound. The identify target command, as its name implies, was used to identify target sounds. During identification the display remained intact; after identification the listener received feedback on accuracy.

The alpha-numeric display presented the same system information alpha-numerically, rather than graphically. The level settings were represented by the name of a dimension and two numbers stating the current range (upper and lower bounds) of that dimension. The two/three dimensions were placed on the same line and the numbers, rather than boxes, changed as a result of a reduction of range. For example, in the two-dimensional database the initial level settings were represented as:

PITCH: LEVELS 1-5   LOUDNESS: LEVELS 1-5.

When the database was played, the level of each dimension used to construct the sound appeared beneath the level settings in this form:

PITCH: LEVELS 1-5   LOUDNESS: LEVELS 1-5

CURRENTLY PLAYING: PITCH: 1 LOUDNESS: 1.

Apparatus. All experimental events were controlled by a general purpose laboratory computer (PDP-11/23). The sounds were output on a 12 bit digital-to-analog converter (Data Translation, model DT-2771) at a sampling rate of 5 kHz, attenuated (Texscan, model SA-50), low-pass filtered at 2.5 kHz (Krohn-Hite, model 3750), and presented binaurally over calibrated, matched headphones (Telephonics, model TDH-50P). Listeners were seated in a soundproof booth (Industrial Acoustics, model 1602A) and a video terminal (Zenith, model WH19) was used to present experimental prompts and to record listener responses. The graphic displays were made with 8 X 10 dot matrix graphic symbols.

Procedure. The experiment was conducted on three consecutive days with each session lasting approximately one hour. In the training session each listener completed a questionnaire to assess his/her computer-related experience. Listeners in each group were trained with one of two graphic displays (abstract or analog). In the training session listeners identified ten sounds in the two-dimensional database. On the second and third day each group used the alpha-numeric display to identify ten sounds in either the two-dimensional or the three-dimensional database. The dimensionality of the database was counterbalanced with the day of testing.

The experimental design contained four independent variables (2X2X2X2 levels), and two dependent variables. The independent variables were graphic display in training session (abstract or analog, between-subjects), dimensionality of database (two- or three-dimensional, within-subjects), experimental trial (first and last five, within-subjects), and order of testing (two- or three-dimensional identifications first, between-subjects). The two

dependent variables were collected on-line: latency of a sound identification and the accuracy of an identification. Identification time was measured (to 1/60 second accuracy).

RESULTS

A normalized accuracy score was computed by comparing the individual's response for each dimension of sound to the actual level used in the target sound's construction. Four scores were computed for each subject by averaging the first and last five trials of the two- and the three-dimensional accuracy scores. A 2X2X2X2 repeated-measures ANOVA was performed on these scores. The main effect of trial, $F(1,36)=8.27$, $p<.01$, and the dimensionality of database by order of testing interaction, $F(1,36)=8.94$, $p<.01$, were significant while all other effects were nonsignificant. The overall accuracy of identification was quite high: 97.75%. Due to the ceiling effect and theoretically uninteresting differences, the accuracy of identification scores will not be discussed further.

Latency scores were collected on-line and represented the elapsed time (in seconds) from the start of a trial to the identification of the target sound. Four scores were computed for each subject by averaging the first and last five trials of the two- and the three-dimensional latency scores. A 2X2X2X2 repeated-measures ANOVA was performed on these scores. The dimensionality of database, $F(1,36)=133.48$, $p<.0001$, the trial, $F(1,36)=42.49$, $p<.0001$, the dimensionality by trial interaction, $F(1,36)=6.54$, $p<.02$, the dimensionality by order interaction, $F(1,36)=52.56$, $p<.0001$, and the graphic display by dimensionality by trial interaction, $F(1,36)=5.97$, $p<.02$, effects were significant while all other effects were nonsignificant. Table 1 illustrates the mean values for latency of

identification in this analysis; the following paragraph summarizes the significant effects.

========================

Insert Table 1 about here

========================

Three-dimensional sounds (mean = 154 sec) took longer to identify than two-dimensional sounds (mean = 105 sec). Identifications took less time on the last five trials (mean = 116 sec) than on the first five trials (mean = 144 sec). Latency for identification of two-dimensional sounds was lower when subjects identified these sounds during Day 3 of the experiment (mean = 84 sec) rather than Day 2 (mean = 126 sec). Likewise, latency for identification of three-dimensional sounds was lower when subjects identified these sounds during Day 3 (mean = 144 sec) than during Day 2 (mean = 164 sec). The dimensionality by trial interaction effect indicated that latency for identification of three-dimensional sounds improved more across trials (means = 173 sec and 136 sec) than latency for two-dimensional sounds (means = 114 sec and 97 sec).

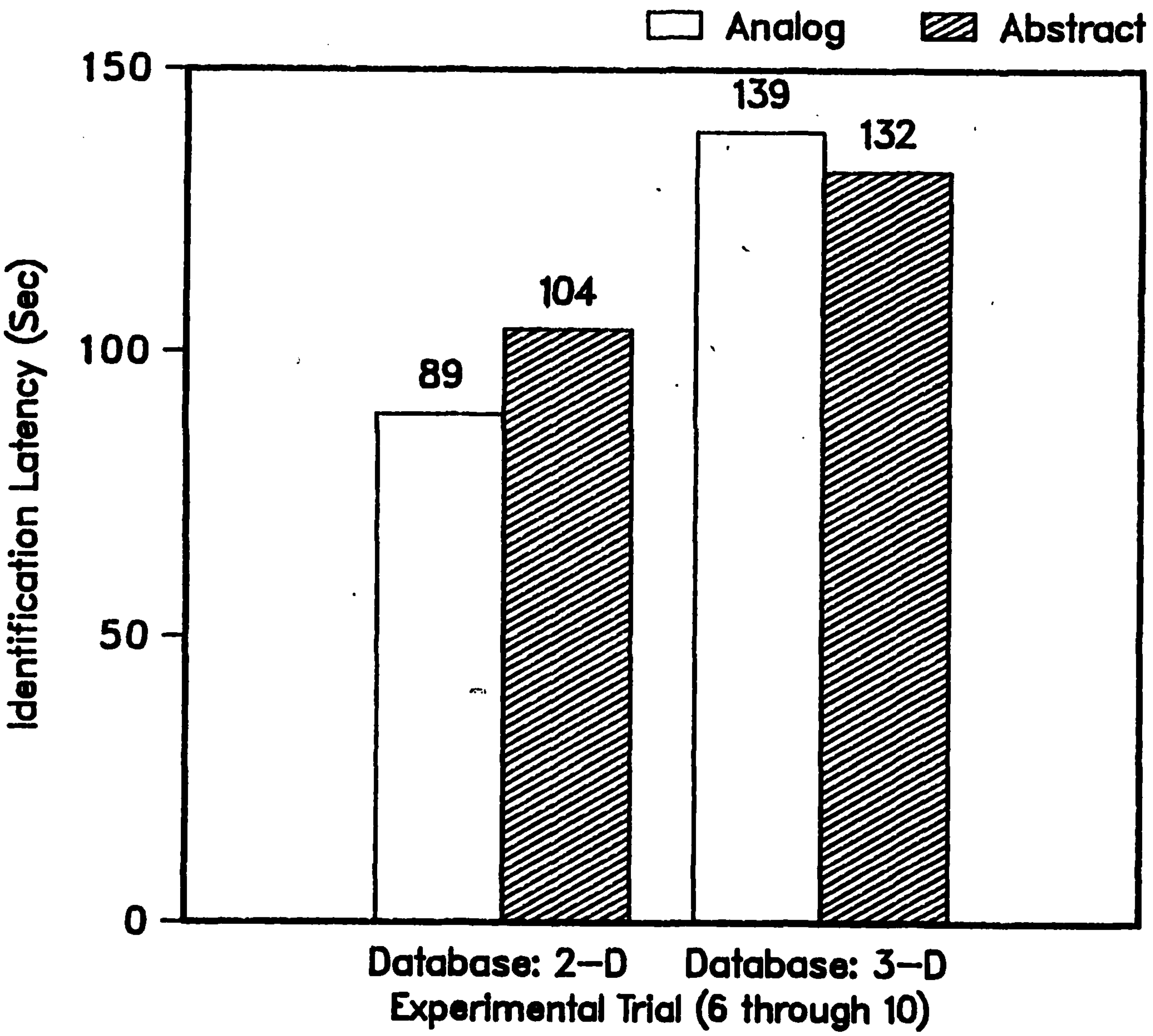
========================

Insert Figure 2 about here

========================

The graphic display by database dimensionality by trial interaction indicated that users trained with the analog graphic display took less time to identify two-dimensional sounds during the last five trials of an experimental session (mean = 89 sec) than users who were trained with the abstract display (mean = 104 sec). A one-tailed t-test indicated that this difference was significant (critical difference at $p < .05$ = 12.64 sec, obtained difference = 14.57

FIGURE 2



Analog
Abstract
150
100
50
0
Identification Latency (Sec)
89
104
139
132
Database: 2–D
Database: 3–D
Experimental Trial (6 through 10)

sec). Although users trained with the abstract graphic display identified three-dimensional sounds during the last five trials in less time (mean = 132 sec) than users trained with the analog graphic display (mean = 139 sec) the difference was not statistically significant. Figure 2 shows the means associated with this effect during the last five trials.

DISCUSSION

The results of Experiment 1 indicate that training with different graphic displays resulted in significant performance differences (latency of indentification) during testing. The interaction between training with a graphic display and the dimensionality of the database was significant, with experimental trials taken into account. Tests for simple effects indicated that users trained with the analog graphic display identified two-dimensional sounds with significantly lower latency during the last five experimental trials than users trained with the abstract graphic display. For three-dimensional sounds the differences were in the predicted direction, but non-significant.

A possible explanation of these results is that training with a graphic display provided users with an internal model of the system (specifically, of the database). When tested with the alpha-numeric display users reasoned about the identification task in terms of the graphic display that they had seen in training. This interpretation is consistent with previous research investigating the role of mental models in performance with a computerized calculator (Halasz and Moran, 1983) and programming languages (Mayer, 1980, 1981). In these studies it was found that training with a model was useful for the solution of novel problems. When faced with a novel problem users

could reason about the device and its internal workings through the model. This allowed users to determine an appropriate course of action.

From this perspective differences in performance could be attributed to the appropriateness of a display for the two- or three-dimensional database. The significantly lower latency scores of users trained with the analog display may have been a result of the one-to-one correspondence between a sound in the database and a [illegible] the screen. This correspondence provided additional [illegible] information to improve the latency of identification [illegible] two-dimensional sounds. Thus, users trained with the analog display developed an internal model more appropriate for the identification of two-dimensional sounds. Similar logic could be applied to explain latency differences for the identification of three-dimensional sounds.

However, the fact that the interaction between training with a graphic display and the dimensionality of the database appeared only as users gained experience with the system (see Fig. 2) is inconsistent with the interpretation that users were reasoning with an internal model based on their graphic display. One would expect these differences to disappear, rather than appear, as users became more familiar with the task. In fact, the results obtained by Halasz and Moran (1983) and Mayer (1980, 1981) indicate that performance differences due to training with a model did disappear as the experimental task became more routine. In these experiments, users were tested on both routine problems (similar to those in training) and novel problems (problems requiring extensions, combinations, or development of new problem-solving strategies) after training with a

model of the device. Although the model supplied users with an aid for the solution of novel problems, once a problem became more routine training with a model did not help in its solution. A second experiment was conducted to investigate whether significant differences could be obtained for a three-dimensional database and to assess the effect of additional experience with the system.

EXPERIMENT 2

Experiment 2 was conducted on five consecutive days. As in Experiment 1, users were trained to search a two-dimensional database with one of two graphic displays on the first day. However, testing continued on four consecutive days and users searched only the three-dimensional database during testing. As in Experiment 1 it was predicted that users trained with the abstract display would identify three-dimensional sounds with lower latency than users trained with the analog display.

A new variable, the inclusion of a practice session prior to experimentation, was included in Experiment 2. In this session users were asked to identify two sounds in the two-dimensional database. During the practice session each user had either 1) an alpha-numeric display (a two-dimensional version of the test display) or 2) the graphic display that the user was trained with. It was predicted that there would be an interaction between the training display (abstract or analog) and the practice display (graphic or alpha-numeric). Daily practice with a graphic display should help users trained with the abstract display but hinder users trained with the analog display.

METHOD

Subjects. Thirty-nine volunteers from the employees of a government research laboratory, aged 19 to 41 years, participated in

the experiment. Three subjects were dropped from the analysis due to failure to complete the task. No listeners reported a history of hearing disorders. Subjects were assigned randomly to graphic display (analog or abstract) and to practice session (appropriate graphic display or modified alpha-numeric display).

Stimuli. The stimuli for Experiment 2 were synthesized on a digital computer using the same algorithms and levels as in Experiment 1. The only change was the removal of the select order command since subjects in Experiment 1 used this command only in an infrequent, exploratory manner.

Apparatus. All experimental events were controlled by a general purpose laboratory computer (PDP-11/70). The sounds were output on a 10 bit digital-to-analog converter (DEC model AR11) at a sampling rate of 5 kHz, attenuated (Hewlett-Packard, model 3500), low-pass filtered at 2.5 kHz (Krohn-Hite, model 3750), and presented binaurally over calibrated, matched headphones (Telephonics, model TDH-50P). Listeners were seated in a soundproof booth (Eckel Industries, model AB200) and a video terminal (Zenith, model WH19) was used to present experimental prompts and to record listener responses.

Procedure. The experiment was conducted on five consecutive days with each session lasting approximately one hour. Before the training session and after the last experimental session each subject completed a questionnaire to assess pre-experimental computer-related experience and post-experimental strategies and impressions.

Subjects were trained with one of two graphic displays and identified ten two-dimensional sounds in the training session (Day 1). On each of the following four days listeners identified two two-dimensional sounds during the practice session and ten

three-dimensional sounds during the experimental session. During the practice session users had either the alpha-numeric display (a two-dimensional version of the test display) or the appropriate graphic display. The experimental design contained four independent variables (2X2X4X2 levels), and two dependent variables. The independent variables were: 1) graphic display in training session (abstract or analog, between-subjects), 2) display in practice session (alpha-numeric or graphic, between-subjects), 3) day of experimental session (one through four, within-subjects), and 4) experimental trial (first and second five, within-subjects). The dependent variable was latency of sound identification and was measured to 1/60 second accuracy.

RESULTS

Eight latency scores were computed for each subject by averaging the time (in sec) for the first and last five trials for each of the experimental sessions. A 2X2X4X2 repeated-measures ANOVA was performed on these scores. The graphic display, $F(1,32)=4.65$, $p<.05$, the graphic display by trial interaction, $F(1,96)=5.79$, $p<.025$, the day of experimental session, $F(3,96)=107.74$, $p<.0001$, the trial, $F(1,32)=54.16$, $p<.0001$, and the day by trial interaction, $F(3,96)=25.26$, $p<.0001$ effects were significant while all other effects were nonsignificant. Table 2 illustrates the mean values for the analysis; the following paragraph summarizes the significant effects.

Users trained with the abstract display identified three-dimensional sounds with lower latency (mean = 98 sec) than users trained with the analog display (mean = 116 sec). A one-tailed t-test indicated that this difference was significant (critical difference at

$p<.05$ = 14.65 sec, obtained difference = 18.33 sec). Users trained with the abstract display identified sounds with lower latency (mean = 128 sec) than users trained with the analog display (mean = 104 sec) during the first five trials averaged for all experimental sessions. A two-tailed t-test indicated that this difference was significant (critical difference at $p<.05$ = 14.79 sec, obtained difference = 23.66 sec). The difference between means for the last five experimental trials was not quite significant (obtained difference = 13.09 sec). Latency improved between experimental sessions (means = 153, 105, 90, and 81 sec) and between the first (mean = 116 sec) and the second five (mean = 98 sec) trials. Within experimental sessions latency improved more during the first five trials (means = 178, 114, 93, and 82 sec) than during the second five trials (means = 128, 97, 87, and 80 sec).

DISCUSSION

The results of Experiment 2 complement the results of Experiment 1. Training with different graphic displays resulted in significantly different latency scores for the identification of three-dimensional sounds. Users trained with the abstract display were able to identify three-dimensional sounds with significantly lower latency than users trained with the analog display. A related finding was that users trained with the analog display performed especially poorly during the first five trials of an experimental session. These results support the conclusion that training with the graphic displays resulted in different mental models of the database system and that differences in the user's mental model of a system can have implications for performance with that system.

That users may have been reasoning about how to use the database system on the basis of an internal model associated with the graphic

displays was put forth as a potential interpretation of the results of Experiment 1. This interpretation is consistent with previous empirical research on mental models (Halasz and Moran, 1983; Mayer, 1980, 1981). However, the results of Experiment 2 are difficult to reconcile with this interpretation. Two aspects of the data are unsupportive: 1) the insignificant interaction between practice display and training display and 2) the subjective reports of users.

========================

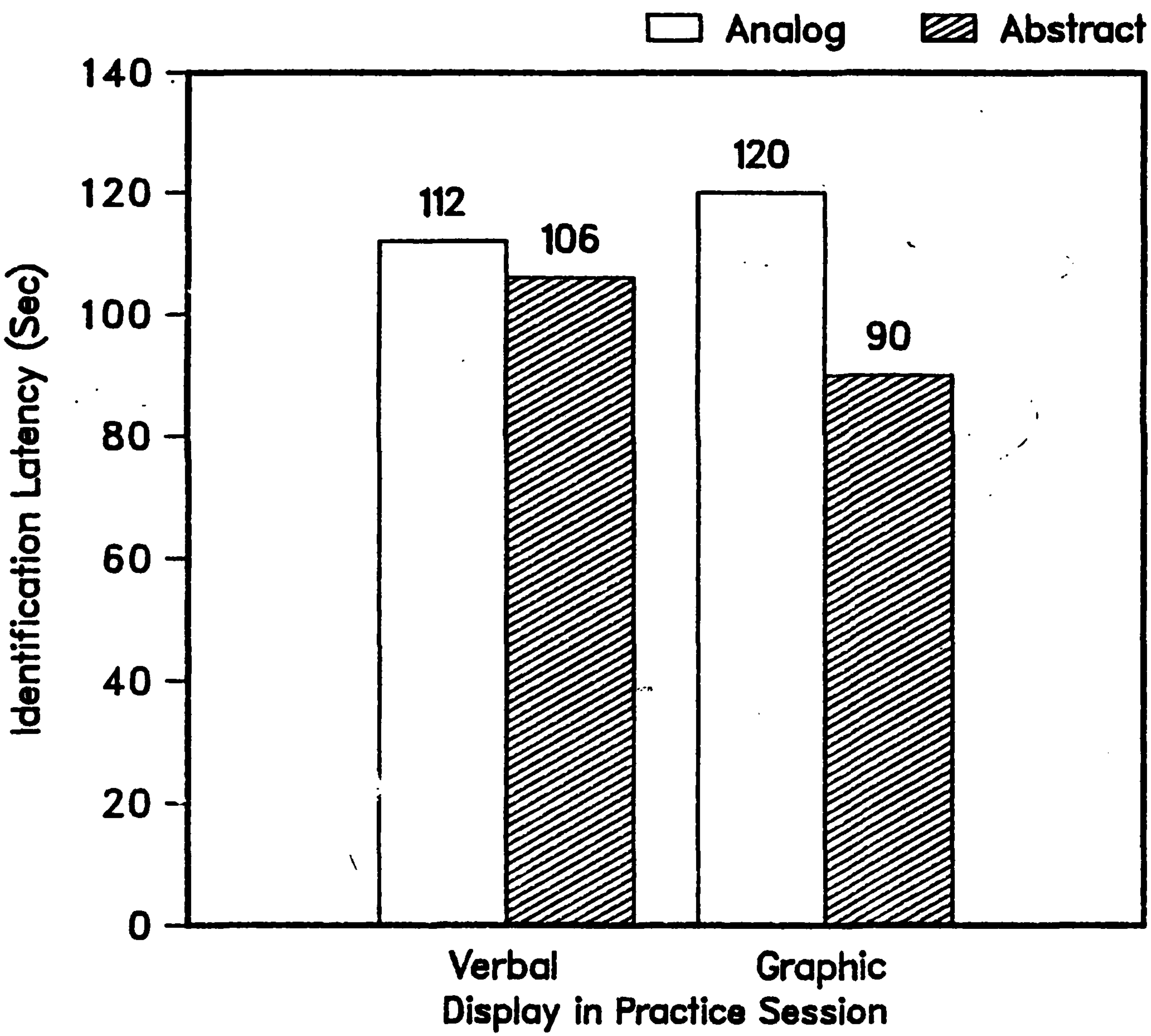
Insert Figure 3 about here

========================

Practice display by training display interaction. If users were reasoning about sound identification on the basis of an internal model then seeing that display (model) every day should have reinforced the internal model, thereby influencing identification times. Figure 3 illustrates that when the practice display was graphic (as opposed to alpha-numeric) the differences were accentuated, and in the predicted direction. However, this effect was non-significant. These results suggest that the user's mental model consisted of more than reasoning on the basis of an internal model associated with the graphic displays.

Subjective reports of users. In a post-experimental questionnaire users answered questions on a scale of 0 to 100 where 0 was labelled Not at All, 50 was labelled Somewhat, and 100 was labelled Extremely. When asked Did you think about the boxes on the screen when you first tried to identify three-dimensional sounds? the average response for all users was 35.7. It is reasonable to assume that if the question had been asked on Day 2 of the experiment the responses would have been somewhat higher. When asked Did you think

FIGURE 3



Analog
Abstract
140
120
100
80
60
40
20
0
Identification Latency (Sec)
112
106
120
90
Verbal
Graphic
Display in Practice Session

about the boxes on the screen after you became practiced at identifying three-dimensional sounds? the average response was 10.5 These results correspond to the subjective reports of users during informal discussion. It can be concluded, at least by the end of Experiment 2, that users were not explicitly reasoning about the experimental task with an internal model based on the graphic displays.

There are other reasons to doubt that the user's mental model consisted solely of [illegible] on the basis of a display-based surrogate mental model. Young (1983, p.42-43) states that "For tasks that require deliberate problem solving ... the surrogate may perhaps be usable as the mental representation on which problem solving is based. But for the more performance-oriented tasks the surrogate seems practically irrelevant." In both experiments experience was found to be a predominant factor determining performance with the database system. In Experiment 1 there were large effects for within-session trial and day of experimental session. Similarly, in Experiment 2 there were significant effects associated with the day of experimental session, the trial, and the display by trial interaction. However, increased experience with the system did not, in general, diminish the effects of training with a graphic display.

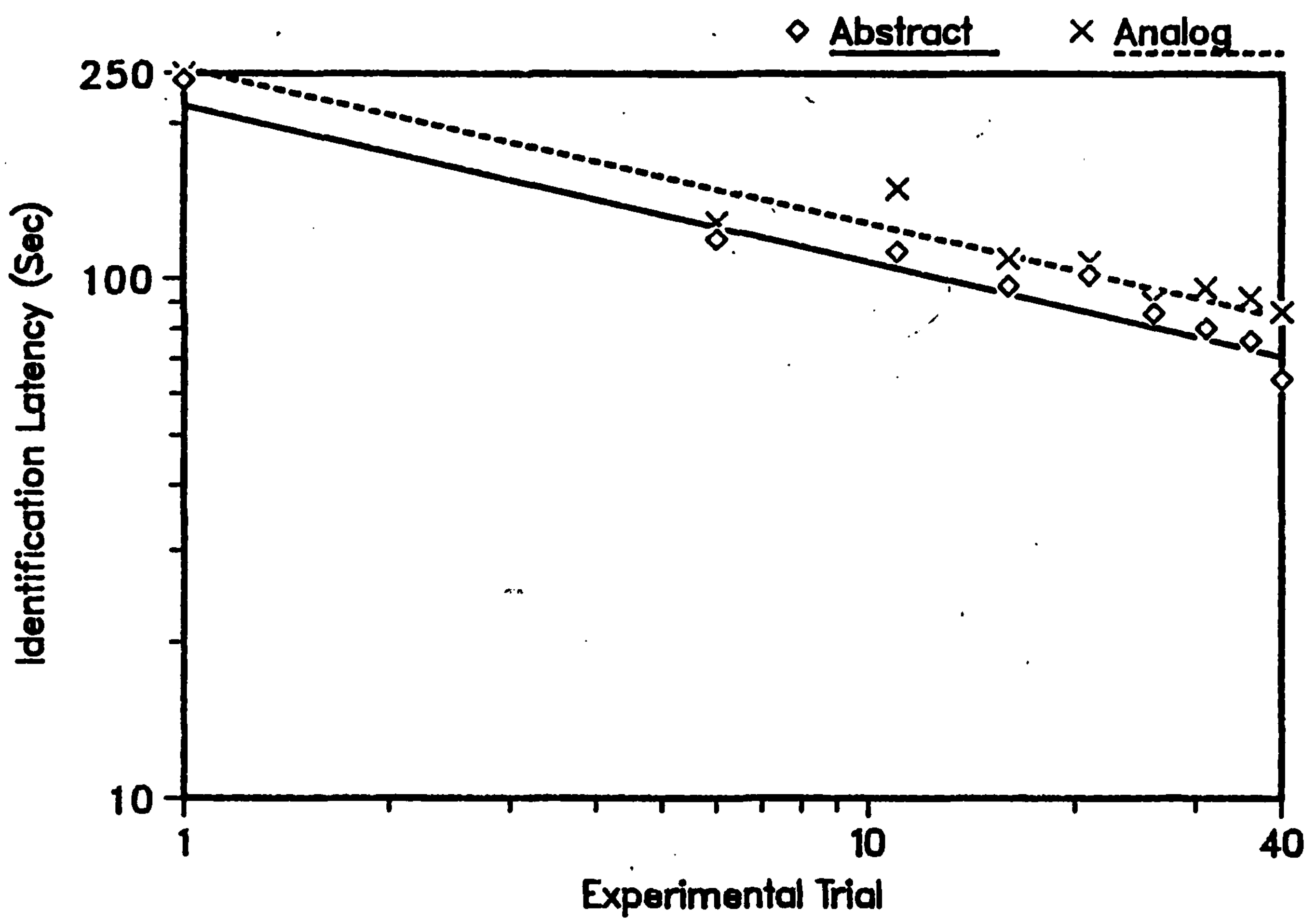
=========================

Insert Figure 4 about here

=========================

This point is particularly clear in the results of Experiment 2. The overall latency of identification (averaged for all users) was lowered from 153 sec on Day 2 to 81 sec on Day 5. This represents a reduction of nearly half. Figure 4 shows a log-log plot of the

FIGURE 4



Abstract
Analog
250
100
10
1
10
40
Identification Latency (Sec)
Experimental Trial

average identification latency for the abstract and analog groups on each trial in the experiment (Although every fifth trial is actually present on the graph, the regression lines were based on all trial values). Thus, the results of the present study complement the findings of Halasz and Moran (1983) and Mayer (1980, 1981) by illustrating that the user's mental model of a system can also influence performance on routine problems.

EXPERIMENT 3

In Experiments 1 and 2 users were trained with either the analog or the abstract graphic display. The results of these experiments indicate that training with a graphic display interacted with the dimensionality of the database: users trained with the analog display identified two-dimensional sounds with decreased latency while users trained with the abstract display identified three-dimensional sounds with decreased latency. It has been argued that training with the abstract display resulted in a mental model of the database system which was more appropriate for the identification of three-dimensional sounds while training with the analog display resulted in a mental model more appropriate for the identification of two-dimensional sounds.

As Stevens and his colleagues (Stevens & Collins, 1980; Stevens, Collins, & Goldin, 1979; Williams, Hollan, & Stevens, 1981) have stressed, reasoning about a complex system may involve the use of several mental models. Williams, Hollan, & Stevens (1981, p. 148) state that they "...consider the use of multiple mental models to be one of the crucial features of human reasoning." Experiment 3 was conducted to investigate whether users could develop multiple mental models of the system. To test this hypothesis users were trained with

both graphic displays, rather than just one. During testing users were shown a graphic display for one trial (similar to the practice trials of Experiment 2) and then tested with the alpha-numeric display for an additional three trials. The user was shown a graphic display to prime a particular mental model.

If users developed separate mental models based on the two graphic displays then seeing a graphic display should invoke the corresponding mental model. This would have implications for subsequent performance with the alpha-numeric display. After priming with the analog display users should perform better on two-dimensional sounds and worse on three-dimensional sounds; after priming with the abstract display users should perform better on three-dimensional sounds and worse on two-dimensional sounds. Thus, based on the results of Experiments 1 and 2, it was predicted that an interaction would occur between priming with a graphic display and the dimensionality of the database.

METHOD

Subjects. Sixteen volunteers from the employees of a government research laboratory, aged 21 to 35 years, participated in the experiment. No listeners reported a history of hearing disorders.

Stimuli. The stimuli for Experiment 3 were synthesized on a digital computer using the same algorithms and levels as in Experiments 1 and 2.

Apparatus. The apparatus was the same as in Experiment 2.

Procedure. The experiment was conducted on three consecutive days with each session lasting approximately one hour. Before the training session and after the last experimental session each subject completed a questionnaire to assess pre-experimental computer-related

experience and post-experimental strategies and impressions.

Subjects were trained with both graphic displays and identified ten two-dimensional sounds in the training session (Day 1). The graphic displays were alternated every other trial; the initial display was randomly determined. During testing (Days 2 and 3) a listener identified a two-dimensional sound with a graphic display and then identified three sounds (either two- or three-dimensional) with the alpha-numeric display. This sequence was repeated four times during each test session. During the first two repetitions users identified two-dimensional sounds. If the analog display was seen on the first repetition then the abstract display was seen on the second (and vice-versa). During the third and fourth repetitions users identified three-dimensional sounds and the presentation order of graphic displays was alternated in a similar fashion. The overall presentation order of graphic displays during testing was counter-balanced: each subject was assigned randomly to one of the sixteen possible combinations.

The experimental design contained four independent variables (3X2X2X2 levels), and two dependent variables. The independent variables were: 1) experimental trial (three sounds identified after a graphic display, within-subjects), 2) graphic display used to prime subjects (abstract or analog, within-subjects), 3) day of experimental session (one or two, within-subjects), and 4) dimensionality of database (two- or three-dimensional, within-subjects). The dependent variable was latency of sound identification and was measured to 1/60 second accuracy.

RESULTS

Twenty-four latency scores were obtained for each subject (twelve

per experimental session). A 3X2X2X2 repeated-measures ANOVA was performed on these scores. The graphic display by trial interaction, $F(2,30)=10.14$, $p<.001$, the day of experimental session, $F(1,15)=17.77$, $p<.001$, the trial, $F(2,30)=14.14$, $p<.001$, and the dimensionality of the database $F(1,15)=50.83$, $p<.001$ effects were significant while all other effects were nonsignificant. The following paragraph summarizes the significant effects.

Two-dimensional sounds (mean = 88 sec) were identified with lower latency than three-dimensional sounds (mean = 117 sec). Experience with the database system again had a significant influence on performance as users improved across trials (means = 115, 99, and 93 sec) and across days (means = 117 and 88). The interaction between graphic display and trial indicated that users took much longer on the first trial after seeing the abstract display (mean = 128) than after the analog display (mean = 103). A one-tailed t-test indicated that this difference was significant (critical difference at $p<.05$ = 11.35 sec, obtained difference = 25.15 sec). As Figure ?? shows, this difference was reversed (but not significantly so) on trials 2 and 3.

DISCUSSION

It was hypothesized that training with both graphic displays would result in multiple mental models of the database. Priming with a graphic display was predicted to invoke one of the two mental models and interact with the dimensionality of the sounds which followed. However, this interaction was not present. Therefore, under the specific circumstances of Experiment 3, users did not develop and employ multiple mental models of the database system.

In retrospect these results may have been predicted from partial results of Experiment 2. In that experiment there were four

experimental groups. The users were trained with one of the two graphic displays. Half of these two groups saw the graphic display that they had been trained with in a daily practice session prior to testing while the other half practiced with the alpha-numeric display. Since only three-dimensional sounds were used in testing it was predicted that daily practice with a graphic display would help users trained with the abstract display but hinder users trained with the analog display. The results did not support the prediction: whether or not the user saw the graphic display that he/she had trained with on each day made no significant difference in performance.

These results, and the results of Experiment 3, strongly suggest that the user's mental model of the system did not consist solely of a mental representation of the system that was based on the graphic displays and used in a deliberate problem-solving manner. This is substantiated by the users' subjective reports in all three experiments. However, the primary results of Experiments 1 and 2 support the mental model hypothesis. It is concluded that the user's mental model of the system was of a more subtle nature than originally predicted. This possibility will be discussed in greater detail in the following section.

GENERAL DISCUSSION

It is often claimed that the user forms a mental model of an interactive computer system which is subsequently used to guide interaction with that system. Despite the popularity of this assumption scant empirical evidence has been provided in its support. The results of the present study indicate that the interface of a computer system with graphics capabilities can contribute to the organization of the user's knowledge about interaction with that

system. That is, interface design can influence the user's mental model of an interactive computing system. Training with an interface containing one of two graphic displays was found to influence performance during testing, when the interface to the system contained an alpha-numeric display. In Experiment 1 users who were trained with the analog graphic display identified two-dimensional sounds in less time than users trained with the abstract graphic display (with increased experience). In Experiment 2 these findings were reversed: users trained with the abstract graphic display identified three-dimensional sounds in less time than those trained with the analog graphic display.

It has often been observed that the representation used in a problem-solving situation can influence the ease of problem solution. This has been noted in the traditional problem-solving literature (e.g., Greeno, 1983) and real-world applications (e.g., Brooke & Duncan, 1981). Brooke & Duncan (1981) describe a study in which display format was altered in a fault-finding task. They conclude that "... modification of the perceptual nature of a display without modification of the basic problem-solving information can affect the speed and efficiency with which a fault in the displayed system is diagnosed" (p. 186). The results of the present study support this conclusion but differ in one aspect: the perceptual modification of the display was not actually present when the data were collected. The observed differences were due to information retained from training with a graphic display: the user's mental model.

The simplest explanation of the present results is that users were reasoning about the identification of sounds on the basis of the graphic displays that they had been trained with. Differences in

performance were due to the appropriateness of a display for the search of the two- or the three-dimensional database. This explanation is similar to what Young (1983) has referred to as reasoning about a device on the basis of a surrogate model. A surrogate model is a simplified, mechanistic account of how a device works. In the present experiments, each graphic display could be considered a surrogate model of the system (specifically, the database).

This interpretation is also similar to what has been referred to in the problem-solving literature as the problem space (Newell and Simon 1972). Halasz and Moran (1983) interpret the results of their research on the mental models of hand-held calculators in this manner, stating that the "problem space is an architectural framework for the knowledge about the possible states of a system, the operations to change the state, and the conditions for the appropriate use of the operations." In the present experiment training with the graphic displays resulted in the formation of different problem spaces for the identification of a sound. At least initially, users trained with the abstract display probably reasoned about operations on vectors (Fig. 1, part b), while users trained with the analog display probably reasoned about operations on a matrix (Fig. 1, part a).

However, as previously mentioned, the results of Experiments 2 and 3 disconfirm the simplistic interpretation that users were reasoning specifically in terms of a graphic display. Also, the users did not feel that they were reasoning on the basis of the graphic displays. One subject's answer to a post-experimental questionaire support this conclusion. When asked the question *When the boxes were not actually on the screen did you imagine that they were? That is,*

did you think about identifying sounds in terms of the graphic displays that you had previously seen? the user replied "I felt comfortable with all three displays after a while. I didn't really think about 'boxes' -- all I thought about were the sounds."

If the users were not specifically reasoning in terms of the graphic displays what is the nature of the user's mental model? It is likely that the cause of performance differences resides in the representation of knowledge in memory. Although information in long-term memory is believed to be stored with one representational system there are two types of short-term or working memories (e.g., Howard, 1983). One working memory represents information with a spatial or visual code while a second working memory represents information with a verbal or linguistic code. This has implications for performance because of the severe capacity and maintenance limitations on information stored in working memory. As Greeno (1983) has noted, the representation of a problem has implications for the ease that analogies can be formed, the information available for reasoning, the efficiency of problem-solving, and planning.

In the present task users would reconstruct a representation of the database (based on the graphic display they had seen in training) using a spatial code in working memory. The observed differences could have been due to the amount of the limited-resource working memory which was required to maintain and reason about the task. Each display was a more or less efficient representation for each database and required more or less effort to maintain the mental representation for reasoning.

However, if the difference was due to different mental representations (spatial codes) in working memory then why did users

claim not to be reasoning in this manner? Anderson's (1982) theory of the acquisition of cognitive skill may shed some light on this apparent discrepancy. The theory draws a major distinction between declarative and procedural knowledge. Declarative knowledge consists of facts about the skill; procedural knowledge consists of how to knowledge. In the context of this experiment, the user must initially think about which command should be used next and whether a potential method is effective or not. However, with increased practice the user integrates these commands into proper sequences and does not have to reason about the task. Declarative knowledge is transformed into procedural knowledge and performance becomes increasingly skilled. When the transformation is complete the individual often looses the ability to verbalize components of the skill. This aspect could account for the user's claim that they no longer reasoned in terms of the graphic display.

Thus, the results of the experiments are interpreted as follows. The graphic displays were interpreted by users as models of the system. The Problem Space theory provides a convenient method of thinking about how the displays influenced initial performance: they provided a different problem space for users to think about interacting with the system. Differences in problem space influenced the users understanding of the function of each command, the internal workings of the database system, and potential methods for using the database system to identify sounds. At least initially, differences in the user's mental model probably included mental imagery (e.g., Shepard, 1978), reasoning by analogy (e.g., Gentner and Gentner, 1983), and beliefs about the the internal workings of the database. As users gained more experience with the database these initial

differences were transformed into differences in procedural knowledge.

CONCLUSIONS

Consideration of the user's mental model in the design of instructional systems is a primary concern, as exemplified in the work of Hollan, Hutchins, and Weitzman (1984) and Williams, Hollan, and Stevens (1981). For novice users a conceptual model which illustrates the important components of a system and how these components interact can facilitate the formation of an appropriate mental model of the system. It allows the user to reason about the system in more familiar or less complex terms.

The results of Experiments 1 and 2 suggest that the interface of any system (instructional or functional) that has graphics capabilities can be interpreted as a model of the system. Users will induce a mental model (Moran, 1981b) through interaction with the system. The results also indicate that the formation of different mental models can have implications for the performance of both novel and routine tasks.

The advent of low-cost computer and graphics technology has resulted in their use in complex man-machine systems. A spatial representation (such as that produced by computer graphics) can have an influence on problem solving. Although computer graphics possess a great potential to improve the man-machine interface, a switch from non-graphic to graphic presentation does not insure this improvement. A system designer must consider the compatibility between a graphic display and the task that the user will be asked to perform. Relatively small differences in design can cause relatively large differences in performance. In the present study graphic representation was shown to influence performance on a task that was

essentially auditory in nature.

ACKNOWLEDGEMENTS

The author would like to thank James H. Howard, Jr. Raja Parasuraman, Frank Yekovich, John J. O'Hare, and Alice J. O'Toole for revisions on earlier drafts. Experiment 1 was conducted at The Catholic University of America in partial fullfillment of the requirements for the degree Doctor of Philosophy and was supported by a contract from the Engineering Psychology Group of the Office of Naval Research. The results were presented at the Mid-Central Ergonomics/Human Factors Conference in Cincinatti on April 12, 1984. Experiments 2 and 3 were conducted at the Naval Underwater Systems Center's CSTL under the sponsorship of the Office of Naval Technology's Postdoctoral Fellowship Program.

REFERENCES

Anderson, J.R. (1982). Acquisition of cognitive skill. Psychological Reveiw, 89, 369-406.

Brooke, J. B., and Duncan, K. D. (1981). Effects of system display format on performance in a fault location task. Ergonomics, 24, 175-189.

Card, S., Moran, T., and Newell, A. (1983). The psychology of human-computer interaction. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Carey, T. (1982). User differences in interface design. Computer, November,14-20.

Carroll, J.M., and Thomas, J.C. (1982). Metaphor and the cognitive representation of computing systems. IEEE Transactions on Systems, Man, and Cybernetics,SMC-12, 107-115.

du Boulay, B., O'Shea, T., and Monk, J. (1981). The black box inside

the glass box: Presenting computing concepts to novices. <u>International Journal of man-machine studies, 14</u>, 237-249.

Gentner, D., and Gentner, D. (1983). Flowing waters or teeming crowds: Mental models of electricity. In D. Gentner and A. L. Stevens (Eds.), <u>Mental Models</u>. Hillsdale, N. J.: Lawrence Erlbaum Associates, 99-129.

Glinert, E. P., and Tanimoto, S. L. (1984). Pict: An Interactive Graphical Programming En[illegible]. [illegible], <u>November</u>, 7-25.

Greeno, J. G. (1983). [illegible]. In D. Gentner and A. L. Stevens (Eds.), <u>Mental Models</u>. Hillsdale, N. J.: Lawrence Erlbaum Associates, 227-252.

Halasz, F., and Moran, T. (1982). Analogy considered harmful. <u>Conference Proceedings: Human Factors In Computing Systems</u>, Gaithersburg, Md., 383-386.

Halasz, F., and Moran, T. (1983). Mental models and problem solving using a calculator. <u>CHI'83 Conference Proceedings: Human Factors In Computing Systems</u>, Boston, Ma., 212-217.

Hollan, J. D., Hutchins, E. L., and Weitzman, L. (1984). Steamer: An interactive inspectable simualation-based training system. <u>The AI Magazine</u>, <u>Summer</u>, 15-27.

Hollnagel, E. and Woods, D.D. (1983). Cognitive systems engineering: New wine in new bottles. <u>International Journal of Man-Machine Studies</u>, <u>18</u>, 583-600.

Howard, D. V. (1983). <u>Cognitive Psychology: Memory, Language, Thought</u>. New York: Macmillan Publishing Co.

Mayer, R.E. (1980). Elaboration techniques for technical text: An Experimental test of the learning strategy hypothesis. <u>Journal of Educational Psychology</u>, <u>72</u>, 770-784.

Mayer, R.E. (1981). The psychology of how novices learn programming. ACM Computing Surveys, 13, 121-141.

Moran, T. (1981a). Guest editor's introduction: An applied psychology of the user. ACM Computing Surveys, 13, 1-12.

Moran, T. (1981b). The command language grammar: A representation for the user interface of interactive computer systems. International Journal of Man-Machine Studies, 15, 3-50.

Newell, A., and Simon, H.A. (1972). Human Problem Solving. Englewood Cliffs, N. J.: Prentice Hall.

Norman, D.A. (1983). Some observations on mental models. In D. Gentner and A. L. Stevens (Eds.), Mental Models. Hillsdale, N. J.: Lawrence Erlbaum Associates, 7-14.

Poulton, E.C. (1966). Engineering psychology. Annual Review of Psychology, 17, 177-200.

Rumelhart, D.E., and Norman, D.A. (1981). Analogical processes in learning. In Anderson, J.R. (Ed.), Cognitive Skills and Their Acquisition. Hillsdale, N.J.: Lawrence Earlbaum Associates, 335-360.

Shepard, R.N. (1978). The mental image. American Psychologist, 33(2), 125-137.

Stevens, A., and Collins, A. (1980). Multiple conceptual models of a complex system. In R.E. Snow, P. Federico, and W.E. Montague (Eds.), Aptitude, Learning, and Instruction, Vol. 2. Hillsdale, N.J.: Lawrence Earlbaum Associates, 177-197.

Stevens, A., Collins, A., and Goldin, S.E. (1979). Misconceptions in students' understanding. International Journal of Man-Machine Studies, 11, 145-156.

Wickens, C.D. (1984). Engineering psychology and human performance.

Columbus, Oh.: Charles E. Merrill Pub. Co.

Williams, M., Hollan, J., and Stevens, A. (1981). An overview of steamer: An advanced computer-assisted instruction system for propulsion engineering. _Behavior Research Methods and Instrumentation_, _13_, 85-90.

Young, R. (1981). The machine inside the machine: users' models of pocket calculators. _International Journal of Man-Machine Studies_, _15_, 51-85.

Young, R. (1983). Surrogates and mappings: Two kinds of conceptual models for interactive devices. (1983). In D. Gentner and A. L. Stevens (Eds.), _Mental Models_. Hillsdale, N. J.: Lawrence Erlbaum Associates, 35-52.

Table 1

Mean values of sound identification latency for graphic display (abstract or analog), dimensionality of database (two- or three-dimensional) and trial (first five and last five) and order of testing (two-dimensional then three-dimensional sounds, or vice-versa)

Display:	Database: 2-D, Trial: 1-5	2-D, 6-10	3-D, 1-5	3-D, 6-10	Averages
Order: 2-D then 3-D sounds					
Abstract	133	124	156	121	134
Analog	143	104	167	132	137
Order: 3-D then 2-D sounds					
Abstract	87	83	197	143	128
Analog	93	73	171	146	121
Averages					
Abstract	110	104	177	132	131
Analog	118	89	169	139	129

Table 2

Mean values of three-dimensional sound identification latency for graphic display (abstract or analog), practice display (graphic or alpha-numeric), day of experimental session (two through five) and within-session trial (first five and last five)

Day of session:	2		3		4		5		
Trial:	1-5	6-10	1-5	6-10	1-5	6-10	1-5	6-10	Averages
Display:									
	Practice display: graphic								
Abstract	136	120	96	87	78	70	67	66	90
Analog	205	136	134	101	103	98	92	93	120
	Practice display: alpha-numeric								
Abstract	184	125	102	95	91	90	82	76	106
Analog	185	130	123	103	96	90	84	82	112
	Averages								
Abstract	160	123	99	91	85	80	75	71	98
Analog	195	133	129	102	100	94	88	88	116

LIST OF FIGURES

END

FILMED

1-86

DTIC